Towards Utilizing Unlabeled Data in Federated Learning: A Survey and Prospective

Yilun Jin[1*] , **Xiguang Wei**[2] , **Yang Liu**[2] and **Qiang Yang**[1,2]
[1]The Hong Kong University of Science and Technology, Hong Kong SAR, China
[2]Webank, Shenzhen, China
yilun.jin@connect.ust.hk, {xiguangwei,yangliu}@webank.com, qyang@cse.ust.hk

Abstract

Federated Learning (FL) proposed in recent years has received significant attention from researchers in that it can bring separate data sources together and build machine learning models in a collaborative but private manner. Yet, in most applications of FL, such as keyboard prediction, labeling data requires virtually no additional efforts, which is not generally the case. In reality, acquiring large-scale labeled datasets can be extremely costly, which motivates research works that exploit unlabeled data to help build machine learning models. However, to the best of our knowledge, few existing works aim to utilize unlabeled data to enhance federated learning, which leaves a potentially promising research topic. In this paper, we identify the need to exploit unlabeled data in FL, and survey possible research fields that can contribute to the goal.

1 Introduction

There should be little doubt that the prosperity of *Artificial Intelligence* (AI) should largely be attributed to the availability of Big Data. As an example, the field of computer vision, where we witnessed numerous advances in deep learning, was significantly boosted with the advent of the comprehensive ImageNet dataset [Deng *et al.*, 2009].

Yet when it comes to applications of AI in real-world scenarios, things are not exactly the case. It is often the case that corporations only possess low-quality, incomplete and insufficient data. To this end, *Federated Learning* [McMahan *et al.*, 2017; Yang *et al.*, 2019; Kairouz *et al.*, 2019] was proposed as an attempt to alleviate such a problem by enabling private collaboration among parties without explicit sharing of data. Up till now, FL has been widely accepted as a new learning scheme and has triggered numerous applications [Hard *et al.*, 2018; Chen *et al.*, 2019; Yang *et al.*, 2018].

Nonetheless, as we observe the existing applications of FL, we find that the majority of them require no additional efforts to label the data. For example, in next-word prediction [Hard *et al.*, 2018], data are automatically labeled through user typing behaviors. Yet in general, raw data collected require manual labeling, which makes it hard to obtain large-scale, high-quality labeled datasets, making the application of FL limited.

We argue that applications of FL are in more pressing need of utilizing unlabeled data than others. On one hand, in cross-device FL [Kairouz *et al.*, 2019], where participants are individual devices, numerous unlabeled data are generated through our interaction with smart devices, such as photos taken, text inputs, and physiological indicators measured by wearables, whose sheer volume makes it impractical to require users to label them. On the other hand, in cross-silo FL where participants are corporations, the data involved are likely to require human expertise, such as finance (risk management, credit evaluation), and medical applications (disease diagnosis, health monitoring). In these cases, it would require significant human intellect and efforts to label the data. In this case, labeling all the data would be costly, and thus makes it necessary to utilize unlabeled data and learn models in a weakly supervised manner.

Nevertheless, compared to other areas, there is relatively little attention paid to this area. While techniques like transfer learning, semi-supervised learning, self-supervised learning and active learning are all popular research topics, we can only observe popularity in federated transfer learning (FTL) [Peng *et al.*, 2020; Liu *et al.*, 2018], while others are relatively ignored.

Consequently, in this paper, we seek to provide a perspective into weakly supervised approaches in federated learning. We first introduce related preliminaries, before identifying motivations that drive us to devote to this problem. Last but not least, we make a prospect into potential scenarios, research topics, as well as challenges. We hope that our efforts can be followed by researchers who come up with concrete solutions to the problem that will contribute to both the academia and the industry.

2 Preliminaries and Related Work

2.1 Federated Learning

Federated Learning, proposed by [McMahan *et al.*, 2017] and extensively surveyed by [Yang *et al.*, 2019; Kairouz *et al.*, 2019], is a machine learning scheme that enables aggrega-

[*] Contact Author

FL setting	ID Space	Feature Space	Label Space
Horizontal Federated Learning (HFL)	Different	Same	Same
Vertical Federated Learning (VFL)	Same/Can be aligned	Different	Different
Federated Transfer Learning (FTL)	Different	(Generally) Different	(Generally) Different

Table 1: FL Categorization According to Data Partition

FL setting	Participants	# Participants	Local Dataset Size	Consistency
Cross-device FL	e.g. phones, IoT devices.	Massive, up to 10^{10} clients	Relatively small	Inconsistent
Cross-silo FL	e.g. corporations, institutes	Up to 10^2.	Relatively large	Consistent

Table 2: FL Categorization According to Type of Participants. The term 'consistency' means the consistency of participants across each round. In cross-device FL, the participants are not always available (e.g. subject to network and battery status, and diurnal-nocturnal changes), making the participants for each round different, and thus 'inconsistent'. On the contrary, cross-silo FL shows much better consistency, as they use dedicated hardware, reliable networks and are much better scheduled.

tion of isolated data in a privacy-preserving manner. Generally speaking there are two major categorization standards proposed by previous surveys, with the first [Yang *et al.*, 2019] focusing on data partitions and the latter [Kairouz *et al.*, 2019] focusing on types of participants. We show the two categorizations in Table 1 and Table 2, respectively.

Existing works on FL have shown significant diversity. There have been research works on federated optimization [McMahan *et al.*, 2017; Li *et al.*, 2020; Wang *et al.*, 2020], federated learning algorithms [Cheng *et al.*, 2019; Li *et al.*, 2019; Shokri and Shmatikov, 2015], privacy mechanisms and attacks [Hitaj *et al.*, 2017; Mohassel and Zhang, 2017; Bonawitz *et al.*, 2017], systems and communication [Bonawitz *et al.*, 2019], etc. However, regarding FL in weakly-supervised scenarios, relatively little attention has been paid to this area.

Existing works on weakly supervised FL mostly fall into federated transfer learning (FTL), with [Peng *et al.*, 2020] and [Liu *et al.*, 2018] proposed unsupervised and supervised FTL, respectively. There are also works tackling federated self-supervised feature learning on texts [Jiang *et al.*, 2019; McMahan *et al.*, 2017] by learning topic models and language models. Regarding other forms of weakly supervised algorithms, such as semi-supervised learning and active learning, we observe little prior arts [Goetz *et al.*, 2019] to the best of our knowledge.

We here discuss two prior works on federated transfer learning. [Liu *et al.*, 2018] tackles the problem of semi-supervised transfer learning between two clients, where the two clients exchange gradients and intermediate results through Homomorphic Encryption (HE). As generally, HE is computationally expensive to perform, the approach may not scale to cross-device FL where maybe millions of participants exist. [Peng *et al.*, 2020] focuses on unsupervised domain adaptation, that uses several source domains held by clients to facilitate classification on one target domain. The work achieves domain adaptation through novel adversarial training techniques and achieved convincing results. Yet, similar to [Liu *et al.*, 2018], this work assumes that the participants are static and constantly available, which also does not scale to the cross-device FL setting.

2.2 Weakly Supervised Learning Algorithms

Transfer Learning

Transfer Learning [Yang *et al.*, 2020] aims to transfer knowledge learned from a source domain to a relevant target domain, probably with fewer labeled samples to train on. Existing popular transfer learning methods include domain adaptation [Long *et al.*, 2014; Long *et al.*, 2015], knowledge distillation [Hinton *et al.*, 2015], and pre-training/fine-tuning [Devlin *et al.*, 2019] etc.

While transfer learning has achieved tremendous success in vision and language modeling, and even triggered interests in FTL, one limitation exists, that a related source domain with abundant data must be found to support transfer learning. In FL, the applications are highly diverse, which makes it hard for every one of them to find a suitable and resourceful source domain.

Semi-supervised Learning

Semi-supervised Learning (SSL) [Zhu, 2005] aims to learn a model under very limited labeled data and also massive unlabeled data. SSL is widely adopted in areas where labels are scarce. In most cases researchers utilize unlabeled data to improve the generalization performance and prevent overfitting caused by small datasets. Popular methods of SSL include generative models [Kingma *et al.*, 2014; Robert *et al.*, 2018], adversarial training [Miyato *et al.*, 2018; Odena, 2016], regularization [Tarvainen and Valpola, 2017], pseudo-labeling [Berthelot *et al.*, 2019], connections between samples [Kipf and Welling, 2016] and multi-view ensemble training [Chen *et al.*, 2018].

Self-supervised Learning

Self-supervised learning, also known as representation learning, aims to extract indicative features from large amounts of data without label supervision. Consequently, common approaches in self-supervised learning utilize the data themselves to provide supervision, trying to capture innate structures within the data. Up till now, self-supervised learning has achieved tremendous success in natural language process (NLP) through large-scale language models [Devlin *et al.*, 2019], and also topic models [Jiang *et al.*, 2019]. Also in the area of vision, self-supervised feature learning is popular, commonly achieved by learning coloriza-

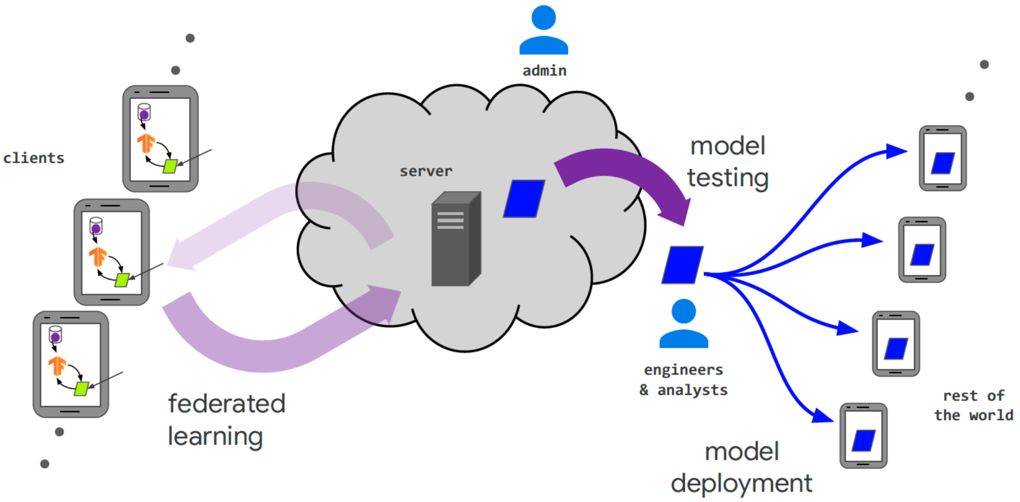

Figure 1: An illustration of cross-device FL. The model is trained through numerous devices and is deployed to all devices throughout the world.

tion, positioning and rotation information [Trinh *et al.*, 2019; Pathak *et al.*, 2016], and has been used for boosting performance in semantic segmentation, clustering, and object detection [Jing and Tian, 2019].

Active Learning

Active Learning aims to train a classifier on datasets with few labeled samples by making as few queries of additional label information as possible. Essentially, active learning aims to find samples that, when labeled, will provide the greatest contribution towards model learning. In existing approaches, active learning is achieved by designing label query algorithms, such as the most uncertain samples [Settles and Craven, 2008], most variance reduction [Schein and Ungar, 2007], etc.

3 Motivations and Advantages

In this section, we identify the motivations that drive us to the problem of FL in weakly supervised settings, and propose advantages that will arise when FL is able to utilize unlabeled samples.

3.1 Expanding Application Scenarios

Existing application scenarios of FL generally work on problems which require little extra effort to label the data. For example, in language modeling [McMahan *et al.*, 2017], labeling is automatically achieved through user typing behaviors. In recommendation [Yang, 2019], the labels are purchase records of users, which also require no extra labor. Yet in most applications, explicit labeling is required, such as object recognition, sentiment analysis, person re-identification, etc.

We also argue that applications of FL face even greater demands in utilizing unlabeled data.

- First, FL imposes strong privacy requirements, which rules out large-scale labeling through outsourcing, which is a common practice in corporations.
- Second, in cross-device FL introduced by [Kairouz *et al.*, 2019], where participants are smart devices, huge amounts of data are generated every day, such as text inputs, images taken, and even physiological indicators measured by wearables. These data are either too large in size to require users to label, or require high-level human expertise (such as sleep monitoring, heartbeats) that few users possess. Consequently, quite often the data generated remain unlabeled.
- Last but not least, in cross-silo FL, where participants are corporations, the data involved often lie within specialized domains, such as finance (risk management, credit evaluation, anti money laundering), or clinical services (medical image diagnosis, object detection and localization). In these domains, the effort required to label the data are generally prohibitive, and therefore we can only afford to label a small proportion of them, instead of the whole dataset.

Consequently, developing algorithms that effectively utilize unlabeled data to enhance training would open up extensive new applications and help build a more vibrant federated AI ecosystem.

3.2 Mitigating Domain Discrepancy

As a challenge identified by many researchers, non-iid data is a prominent issue in FL, and there have also been works to study such a challenge [Li *et al.*, 2020]. Generally speaking, non-iid data pose two challenges to FL. On one hand, the data owned by different parties inevitably differ in their distribution, causing difficulties in model learning. On the other hand, domain discrepancy also exists between training and testing. Chances are that the data used to train a federated model differs a lot to those owned by certain users, making the model ineffective for them. In fact, a recent empirical study [Yu *et al.*, 2020] demonstrated that, federated language models can be less accurate than a considerable proportion (as much as 20%) of local models trained using data from individual parties, whose data distributions differ a lot from the global distribution.

Utilizing large-scale unlabeled data, correspondingly, is able to mitigate the problem of non-iid data. Intuitively, by viewing a sufficiently large unlabeled dataset, one can get a much better understanding of the data distribution than using only a small labeled dataset alone. For example, unlabeled data can be used to train generative models that provide additional information about the data's prior distribution $p(x)$, thus filtering out the domain-specific features [Kingma *et al.*, 2014; Robert *et al.*, 2018]. In addition, domain adaptation that minimizes domain discrepancies can also be used on unlabeled data [Peng *et al.*, 2020], such that domain invariant representations can be learned. Last but not least, advances in disentangled representations [Siddharth *et al.*, 2017] can also contribute to domain invariant models by disentangling domain-specific features from domain invariant ones.

3.3 Enhancing Robustness

Robustness means that a model would be resilient to small variations, such as outliers and small perturbations of inputs, which is appealing in most machine learning applications. By utilizing unlabeled data to regularize the model, robustness can be achieved. For example, sensitivity towards small perturbations can be alleviated if we regularize the model to produce consistent outputs in the neighborhood of each data point. It would not be possible if only a few labeled samples are available, as they only represent a small subset over the data distribution. In addition, reliance on specific data points can be alleviated if more unlabeled data can be used to prevent overfitting on a few labeled samples.

Robustness in FL also implies attractive outcomes. On one hand, when participants of FL have a rather limited amount of data, the local trainings are likely to be noisy, and local models prone to overfitting. By utilizing available unlabeled data for regularization, local overfitting can be alleviated and therefore, a better global model can be reached. On the other hand, robustness implies resilience towards modification of the dataset, which is favorable towards private and secure machine learning models. For example, robustness against small perturbations would lead to resistance over data poisoning attacks, such as adversarial examples [Goodfellow *et al.*, 2014]. As another example, as shown in [Shokri *et al.*, 2017], membership inference attacks are closely related to overfitting, and the more overfitting the model is, the more prone it is towards membership inference attacks (as the model is more likely to behave differently on samples that are used to train the model). Consequently, robustness in FL can also lead to appealing properties in security.

4 Potential Topics and Challenges

In this section we introduce potential settings and topics, both in research and applications, that may contribute to better FL algorithms, and also potential challenges that may arise.

4.1 Transfer Learning

Existing solutions enabling FTL have been highly sophisticated [Liu *et al.*, 2018; Peng *et al.*, 2020]. We here identify several potential topics regarding FTL.

- **Versatile Source Domains and Datasets.** As FL should support a wide range of applications, to enable FTL, it is important that adequate source domains and datasets are chosen, otherwise negative transfer [Cao *et al.*, 2010] may happen. It is thus important in practice that adequate source domains must be chosen to enable FTL applications. Alternatively, it is always welcomed to develop versatile datasets that transfer to multiple domains.
- **Realistic Federated Datasets.** FL features non-iid data held by different participants, as determined by location, population, etc, and a realistic federated dataset that accurately replicates such domain discrepancies would be necessary for evaluating FTL or even broader FL algorithms. Up till now, existing FTL evaluations use artificial datasets created by manipulating existing benchmarks, which may not accurately capture real-world domain discrepancies featured by FL.
- **FTL in cross-device FL.** Existing solutions on FTL work on relatively few participants, e.g. several, or tens, with each of them holding relatively large data, and are always available throughout the training [Peng *et al.*, 2020]. Yet, in cross-device FL, participants are much larger in size, inconsistent for each round of training, and each of them may hold much smaller amounts of data, as shown in Table 2. It is thus relatively unknown how FTL can work in the cross-device FL setting, which also shows significant domain discrepancy [Yu *et al.*, 2020].

4.2 Semi-supervised Learning

Semi-supervised setting in FL has received little attention, which leaves a promising potential topic, as semi-supervised learning can work on almost all types of data. For example, in medical image classification, obtaining fully annotated training datasets may not be possible, where we can resort to federated semi-supervised learning to solve the problem. As another example, it is also costly to obtain fully annotated data in financial applications, where collaborators such as banks, insurance companies would jointly train their model in a semi-supervised manner. We here point out several potential challenges that need to be resolved in this topic.

- **Privacy Requirements.** In certain semi-supervised learning algorithms, connections between samples are leveraged to infer or 'propagate' labels towards unlabeled samples [Kipf and Welling, 2016]. In these approaches, it is important that privacy requirements are not breached when we leverage these connections. There are also algorithms that involve generative models, which are capable of generating artificial samples [Robert *et al.*, 2018; Springenberg, 2015]. Whether such artificial samples are breaking the privacy requirements remains an important challenge that is yet to be resolved.
- **Domain Discrepancy** Non-iid data always pose significant challenges in FL. In the case of semi-supervised

learning, [Oliver *et al.*, 2018] showed that when labeled data and unlabeled data belong to different domains (i.e. domains that show significant discrepancy), semi-supervised learning algorithms will significantly degrade in performance. Thus, semi-supervised learning methods in FL must be combined with techniques that tackle with domain discrepancy.

- **Extension to VFL** Existing studies on semi-supervised learning mainly fits in with the HFL setting, where the unlabeled data are shown intact. However, when it comes to VFL, where the data samples themselves are fragmented and cannot be brought together, more sophisticated protocols should be designed.
- **Relationship between robustness and security** As mentioned before, model robustness (e.g. robustness to perturbations, outliers) are intuitively related with defense against attacks, such as adversarial attacks and membership inference attacks. As various regularization techniques are involved in semi-supervised learning, it is interesting to study, both empirically and theoretically how such regularization and robustness will contribute towards model security.

4.3 Self-supervised Learning

One significant doubt on self-supervised learning in FL is that, it may depend strongly on the data domain and the downstream task it is used for. For example, while self-supervised language modeling is competitive in a wide range of tasks, self-supervised learning in vision is not the case. As shown in [Goyal *et al.*, 2019], self-supervised learning is competitive in object detection, but outperformed by supervised pretraining significantly in various classification tasks. Consequently, although self-supervised learning is a natural idea in FL [McMahan *et al.*, 2017], whether it enables wider application may be doubtful and depend heavily on the specific application.

4.4 Active Learning

Active learning seems a natural idea that can be well combined with FL. For example, in cross-device FL, the model holder may ask certain users to label several examples which are then used for training, acting in a crowd-sourcing manner. In cross-silo FL, an institute may identify several difficult examples during training, and ask its experts to label it to facilitate training.

A key challenge that needs to be solved is how to identify data samples that contribute most to training and should be queried. In federated learning, neither the coordinator or the training server can directly observe raw data. Instead they can only observe batched, and in some cases even protected (e.g. via differential privacy) or encrypted intermediate results. Consequently, identifying individual data samples that may contribute most to training is not straightforward.

5 Conclusion

In this paper we identify a potentially important topic in federated learning: utilizing unlabeled data for weakly supervised federated training. We introduce existing methods that effectively leverage unlabeled data for training models, and point out motivating advantages that arise if unlabeled data can be incorporated for weakly-supervised training. Finally, we make a prospect into potential topics, application scenarios and challenges that come along weakly supervised learning in FL. We hope that this paper can lead to more attempts in more effective utilization of data, better learning algorithms, and a more diverse federated ecosystem featuring a wider range of applications.

References

[Berthelot *et al.*, 2019] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 5049–5059. Curran Associates, Inc., 2019.

[Bonawitz *et al.*, 2017] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1175–1191. ACM, 2017.

[Bonawitz *et al.*, 2019] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konecny, Stefano Mazzocchi, H Brendan McMahan, et al. Towards federated learning at scale: System design. *arXiv preprint arXiv:1902.01046*, 2019.

[Cao *et al.*, 2010] Bin Cao, Sinno Jialin Pan, Yu Zhang, Dit-Yan Yeung, and Qiang Yang. Adaptive transfer learning. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[Chen *et al.*, 2018] Dong-Dong Chen, Wei Wang, Wei Gao, and Zhi-Hua Zhou. Tri-net for semi-supervised deep learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 2014–2020. AAAI Press, 2018.

[Chen *et al.*, 2019] Mingqing Chen, Rajiv Mathews, Tom Ouyang, and Françoise Beaufays. Federated learning of out-of-vocabulary words. *arXiv preprint arXiv:1903.10635*, 2019.

[Cheng *et al.*, 2019] Kewei Cheng, Tao Fan, Yilun Jin, Yang Liu, Tianjian Chen, and Qiang Yang. Secureboost: A lossless federated learning framework. *arXiv preprint arXiv:1901.08755*, 2019.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[Devlin *et al.*, 2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[Goetz *et al.*, 2019] Jack Goetz, Kshitiz Malik, Duc Bui, Seungwhan Moon, Honglei Liu, and Anuj Kumar. Active federated learning. *arXiv preprint arXiv:1909.12641*, 2019.

[Goodfellow *et al.*, 2014] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[Goyal *et al.*, 2019] Priya Goyal, Dhruv Mahajan, Abhinav Gupta, and Ishan Misra. Scaling and benchmarking self-supervised visual representation learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6391–6400, 2019.

[Hard *et al.*, 2018] Andrew Hard, Kanishka Rao, Rajiv Mathews, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon, and Daniel Ramage. Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*, 2018.

[Hinton *et al.*, 2015] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[Hitaj *et al.*, 2017] Briland Hitaj, Giuseppe Ateniese, and Fernando Perez-Cruz. Deep models under the gan: information leakage from collaborative deep learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 603–618. ACM, 2017.

[Jiang *et al.*, 2019] Di Jiang, Yuanfeng Song, Yongxin Tong, Xueyang Wu, Weiwei Zhao, Qian Xu, and Qiang Yang. Federated topic modeling. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1071–1080, 2019.

[Jing and Tian, 2019] Longlong Jing and Yingli Tian. Self-supervised visual feature learning with deep neural networks: A survey. *arXiv preprint arXiv:1902.06162*, 2019.

[Kairouz *et al.*, 2019] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.

[Kingma *et al.*, 2014] Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In *Advances in neural information processing systems*, pages 3581–3589, 2014.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Li *et al.*, 2019] Qinbin Li, Zeyi Wen, and Bingsheng He. Practical federated gradient boosting decision trees. *arXiv preprint arXiv:1911.04206*, 2019.

[Li *et al.*, 2020] Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, and Zhihua Zhang. On the convergence of fedavg on non-iid data. In *International Conference on Learning Representations*, 2020.

[Liu *et al.*, 2018] Yang Liu, Tianjian Chen, and Qiang Yang. Secure federated transfer learning. *arXiv preprint arXiv:1812.03337*, 2018.

[Long *et al.*, 2014] Mingsheng Long, Jianmin Wang, Jiaguang Sun, and S Yu Philip. Domain invariant transfer kernel learning. *IEEE Transactions on Knowledge and Data Engineering*, 27(6):1519–1532, 2014.

[Long *et al.*, 2015] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*, pages 97–105, 2015.

[McMahan *et al.*, 2017] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pages 1273–1282, 2017.

[Miyato *et al.*, 2018] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1979–1993, 2018.

[Mohassel and Zhang, 2017] Payman Mohassel and Yupeng Zhang. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 19–38. IEEE, 2017.

[Odena, 2016] Augustus Odena. Semi-supervised learning with generative adversarial networks. *arXiv preprint arXiv:1606.01583*, 2016.

[Oliver *et al.*, 2018] Avital Oliver, Augustus Odena, Colin A Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 3235–3246. Curran Associates, Inc., 2018.

[Pathak *et al.*, 2016] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

[Peng *et al.*, 2020] Xingchao Peng, Zijun Huang, Yizhe Zhu, and Kate Saenko. Federated adversarial domain adaptation. In *International Conference on Learning Representations*, 2020.

[Robert *et al.*, 2018] Thomas Robert, Nicolas Thome, and Matthieu Cord. Hybridnet: Classification and reconstruction cooperation for semi-supervised learning. In *Pro-*

ceedings of the European Conference on Computer Vision (ECCV), pages 153–169, 2018.

[Schein and Ungar, 2007] Andrew I Schein and Lyle H Ungar. Active learning for logistic regression: an evaluation. *Machine Learning*, 68(3):235–265, 2007.

[Settles and Craven, 2008] Burr Settles and Mark Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 1070–1079, 2008.

[Shokri and Shmatikov, 2015] Reza Shokri and Vitaly Shmatikov. Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, pages 1310–1321. ACM, 2015.

[Shokri *et al.*, 2017] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3–18. IEEE, 2017.

[Siddharth *et al.*, 2017] Narayanaswamy Siddharth, Brooks Paige, Jan-Willem Van de Meent, Alban Desmaison, Noah Goodman, Pushmeet Kohli, Frank Wood, and Philip Torr. Learning disentangled representations with semi-supervised deep generative models. In *Advances in Neural Information Processing Systems*, pages 5925–5935, 2017.

[Springenberg, 2015] Jost Tobias Springenberg. Unsupervised and semi-supervised learning with categorical generative adversarial networks. *arXiv preprint arXiv:1511.06390*, 2015.

[Tarvainen and Valpola, 2017] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in neural information processing systems*, pages 1195–1204, 2017.

[Trinh *et al.*, 2019] Trieu H Trinh, Minh-Thang Luong, and Quoc V Le. Selfie: Self-supervised pretraining for image embedding. *arXiv preprint arXiv:1906.02940*, 2019.

[Wang *et al.*, 2020] Hongyi Wang, Mikhail Yurochkin, Yuekai Sun, Dimitris Papailiopoulos, and Yasaman Khazaeni. Federated learning with matched averaging. In *International Conference on Learning Representations*, 2020.

[Yang *et al.*, 2018] Timothy Yang, Galen Andrew, Hubert Eichner, Haicheng Sun, Wei Li, Nicholas Kong, Daniel Ramage, and Françoise Beaufays. Applied federated learning: Improving google keyboard query suggestions. *arXiv preprint arXiv:1812.02903*, 2018.

[Yang *et al.*, 2019] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):12, 2019.

[Yang *et al.*, 2020] Qiang Yang, Yu Zhang, Wenyuan Dai, and Sinno Jialin Pan. *Transfer learning*. Cambridge University Press, 2020.

[Yang, 2019] Qiang Yang. Federated recommendation systems. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 1–1. IEEE, 2019.

[Yu *et al.*, 2020] Tao Yu, Eugene Bagdasaryan, and Vitaly Shmatikov. Salvaging federated learning by local adaptation. *arXiv preprint arXiv:2002.04758*, 2020.

[Zhu, 2005] Xiaojin Jerry Zhu. Semi-supervised learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2005.